

استعمال الطرائق الحصينة لتقدير معالم نموذج الاقتران لجداول التوافق

بوجود خلايا شاذة مع التطبيق العملي

أ.د. دجلة ابراهيم مهدي

طالب الماجستير. أحمد رزاق عبد رمضان (*)

كلية الإدارة والاقتصاد / جامعة بغداد

المبحث الاول : منهجية البحث

1-1 المستخلص

تناولنا في هذه البحث تقدير معالم نموذج الاقتران في حالة وجود خلايا شاذة حيث تعد وسيلة مهمة لتحليل البيانات لمعرفة اي من طرائق التقدير الحصينة هي الأفضل والأكثر كفاءة. عند استعمال طريقة تقدير (M) باستعمال الدوال (Huber, Tukey, Andrews). وعند استعمال طريقة (Kuhnt) للكشف عن الشواذ في جداول التوافق لبيانات حقيقية تخص ظاهرة البطالة لعينة قوامها (1834) عاطل في محافظة واسط وجدت ثلاث خلايا شاذة اي ما يعادل نسبة (10%) من البيانات وعليه اتضح أن مقدرات (M) الحصينة دالة (Andrews) هي الأفضل بين الدوال المستخدمة وهذا ما أثبتته معيار المفاضلة (متوسط مربعات الخطأ MSE), وأثبتت نتائج التجربة أن نسبة العاطلين عن العمل اغلبهم من الشباب حملة الشهادات الدنيا وهو مطابق للواقع وذلك لقلّة فرص العمل لديهم.

Abstract

In this research estimate the parameters of the model association in the case of the existence outliers cells are an important way to analyze the data to see any of the estimation methods classical and robustness is the best and most efficient when using M-Estimators using the functions (Huber, Tukey, Andrews) When using method (Kuhnt) for the detection of outliers in the Contingency Tables to the real data concerning the phenomenon of unemployment for a sample of (1834) unemployed, in Wasit City, and found three of outliers cells is equivalent to the proportion (10%) of the data, it turns out that robust M-Estimators, the function (Andrews) is the best among the functions used and this is proven by the standard differentiation (Mean Square Error MSE),

(*) جزء مستل من رسالة ماجستير للباحث الثاني.

The Results of the experiment proved that the proportion of the unemployed most of them from young graduates a low and this is identical to the reality, because do not have employment opportunities.

Introduction

2-1 المقدمة

نتيجة للتطور الذي بدأ في ستينيات القرن المنصرم تطورت أساليب تحليل البيانات المصنفة واستمر التطور الى هذا اليوم مرافقا للتطور الذي حصل في مختلف المجالات ومنها الحاسوب ومن هذا كان لعلم الإحصاء دورا مهما في جعل البيانات المصنفة مادة أساسية لإجراء التجارب والبحوث وخصوصا في مجالات الصحة والعلوم النفسية و الإنسانية فضلا عن استخدام البيانات العنقودية^[3]

ويعتبر الباحث كارل بيرسن (Karl Pearson) أول من قام بتحليل البيانات المصنفة في عام 1900 وقد اعتبرت هذه الخطوة تطورا نسبيا في نماذج الاستجابة حتى وصل التطور الذي حدث في ستينيات القرن الماضي وعليه كان لابد من تطور الأدوات والأساليب التي تختص بالتحليل إذ ظهرت طرائق جديدة لتحويل هذه البيانات لا تعتمد على النظرة التقليدية باستخراج العلاقة باستخدام مربع كاي لتحديد درجة العلاقة. بل ظهرت أشكال أخرى منها التحليل المتناظر والتحليل اللوغاريتمي , أما التحليل المتناظر يعتمد بصورة خاصة على عرض البيانات بشكل بياني مشترك للصفوف والأعمدة وذو أبعاد قليلة تمكن القارئ من فهم العلاقات في مصفوفة البيانات من خلال تحليل الشكل البياني هذا^[2] أما التحليل اللوغاريتمي الخطي فيعتمد على أنموذج رياضي محدد بمجموعة من القيود والافتراضات التي ليس من السهل توفرها لكل مجموعة بيانات .

وتصنف المتغيرات في جداول توافق اما مستمرة او متقطعة فعندما تكون هناك علاقة متداخلة بين بيانات متعلقة بمتغيرات مستمرة يمكن تحليل هذه العلاقة في اغلب الاحيان بواسطة النماذج الخطية . اما اذا كانت هذه العلاقة متعلقة بمتغيرات متقطعة فيمكن تحليل هذه العلاقة بواسطة جداول توافق متعددة الاتجاهات , إذ ان لكل نوع من البيانات ولكل شكل من هذه الأشكال فرضياته وطرائقه في التحليل . وعلى هذا الأساس سيتم البحث على دراسة جداول التوافق بدرجاتها المختلفة و أنواعها والنماذج اللوغاريتمية لها .

3-1 مشكلة البحث:

غالبا ما تواجه الباحثين وخصوصا في عملية جمع البيانات احتواء تلك البيانات على قيم شاذة تنشأ نتيجة أخطاء الاستبيان او عدم الإفصاح بشكل الصحيح عن المعلومات بالنسبة للعينة المستبانه وخصوصا في تلك البيانات التي توصف على أساس أنها بيانات مصنفة في جداول التوافق والتي تشكل القيم الشاذة فيها خلايا شاذة يكون وجودها عدم مقدره الطرائق التقليدية من إيجاد التقديرات الصحيحة لتلك الخلايا وبالتالي عدم كفاءة تلك الطرائق مثلا ظاهرة البطالة احد أهم تلك المتغيرات التي غالبا ما يصاحبها وجود مثل تلك الخلايا الشاذة وان ظاهرة البطالة من الظواهر الخطيرة وان اغلب بيانات البطالة تكون عادة مصنفة وكثير ما تحتوي على الخلايا الشاذة وذلك للأخطاء الحاصلة في تسجيل البيانات بعض هذه الأخطاء تؤدي الى تشكيل خلايا شاذة حيث أن اغلب العاملين بأجور او كسبة يسجلون على أنهم عاطلين خلال عملية جمع البيانات .

اذن من الواجب إيجاد طرائق أخرى لمعالجة تلك المشاكل والكشف عن تلك الخلايا الشاذة وهناك أمثلة كثيرة في واقعنا العملي تظهر من خلالها الخلايا شاذة مثلا ربة البيت أو سائق التاكسي الخ

Purpose of Search

4-1 هدف البحث

يهدف البحث إلى تقدير معلمات نموذج الاقتران لجدول التوافق ذات الرتبة $(R * C)$ في حالة وجود خلايا شاذة باستعمال الطرائق الحصينة ومن ثم المقارنة بين طرق الحصينة من خلال تطبيقها على بيانات ظاهرة البطالة في محافظة واسط باستعمال جداول التوافق ذات الرتبة $(R * C)$ بدراسة متغيرين مهمين لهذه الظاهرة والمتمثلان بمتغير التحصيل الدراسي و متغير العمر للشخص العاطل .

المبحث الثاني : الجانب النظري**1-2 المقدمة:**

ان اغلب البحوث التي عرضت تحليل البيانات المصنفة خصوصاً تلك الموضوعية على شكل جداول توافق استخدام النماذج اللوغاريتمية الخطية وتضمنت اختبار الفرضيات الملائمة منها وجود العلاقة بين المتغيرات من عدمها وتحديد النماذج المناسبة لتمثيل تلك البيانات وأساليب تقدير معالم تلك النماذج بمختلف الطرائق المعلمية واللامعلمية إلا ان هذه النماذج مثل سائر نماذج الانحدار تخضع لمجموعة من الفرضيات والشروط لكي يتم الحصول على مقدرات كفوءة خصوصاً في حاله عدم وجود قيم شاذة (Outliers) إلا ان تواجدها نتيجة الأخطاء تسجيل البيانات او المشاهدات يؤدي الى مقدرات غير كفوءة التي تم تقديرها بالطرائق التقليدية يصبح من الضروري استخدام الطرائق الحصينة كضرورة ملحه للحصول على مقدرات كفوءة لا تتأثر بالشواذ لذا سيتم في هذا البحث عرض جداول التوافق وتصنيفها وعرض النماذج اللوغاريتمية لجدول توافق ذات اتجاهين وكذلك نماذج الاقتران ومفهوم الشواذ في جداول التوافق وأخيرا تقدير معالم الأنموذج اللوغاريتمي بأحد أهم الطرائق الحصينة باستخدام أسلوب M وفقاً لدالة Andrews ,Tukeys ,Huber.

Contingency tables**5-2 جداول التوافق [2] [3] [5] [12] [13] [16] [27]**

يتم تكوين جداول التوافق عندما تعبر او تمثل البيانات المصنفة على شكل أصناف او مجاميع او مستويات تخص المتغيرات تحت الدراسة وقد يكون الجدول لمتغيرين بمستويات مختلفة او أكثر وهذه المتغيرات قد تكون نوعيه مثل (لون العين , لون الشعر) او تكون كمية مقاسه مثل (الطول , الوزن.....), وتعرف جداول التوافق على انها "ترانيب او تصنيفات من التكرارات موضوعه في جدول يضم متغيرين او أكثر وكل متغير يقسم الى مستويات او تصنيفات عدة [19]

Association**6-2 نماذج الاقتران [12] [18] [19] [20] [21] [22] [28]****Model**

تعد نماذج الاقتران نوعاً من أنواع النماذج اللوغاريتمية او يمكن عده حاله خاصة من النماذج اللوغاريتمية الخطية (Log Linear) والتي يعود الفضل في اكتشافها للباحث (Goodman) في عام 1979 وان نماذج الاقتران هذه تقترض نفس تلك القيود والشروط التي تقترضها النماذج اللوغاريتمية , والهدف من نموذج الاقتران هو قياس قوة العلاقة بين اثنين او أكثر من المتغيرات المرتبة فلو توفر جدول توافق ذي الاتجاهين (r) من الصفوف و(c) من الأعمدة وبحاله يمكن معها الترتيب مستويات الصفوف و الأعمدة (v_j,u_i) على التوالي وبشرط ان (u₁<u₂<u₃<.....<u_r) بالنسبة للصفوف و (v₁<v₂<v₃<.....<v_c).

اذ ان الصيغة العامة لأنموذج الاقتران يعبر عنها وفقاً كالاتي :

$$\log m_{ij} = \mu + \alpha_i^R + \beta_j^C + \gamma U_i V_j \quad \dots (1)$$

اذ ان :

μ : المتوسط العام .

β_j^C : متجه معلمات الصفوف

α_i^R : متجه معلمات الأعمدة .

γ : معلمه قياس العلاقة بين المتغير الاول X والمتغير الثاني Y اي الصفوف

و الأعمدة على التوالي.

وهناك عده أنواع لنماذج الاقتران وبحسب طريقة وضع الدرجات (Score) واهم هذه الأنواع:

1-6-2 نموذج الاقتران الصفري Null Association وهو النموذج الذي يفتر الى التفاعل او العلاقة بين المتغيرات وتكون صيغته كالاتي :

$$\log m_{ij} = \mu + \alpha_i^R + \beta_j^C \dots (2)$$

2-6-2 نموذج الاقتران المنتظم Uniform Association Model يسمى نموذج الاقتران بالمنتظم عندما تكون درجات (الرتب) لمستويات الصفوف و الأعمدة مساوية لرقم الصف والعمود او بمعنى آخر عندما تكون $u_i=i$ و $v_j=j$ وتكون الرتب هي أعداد صحيحة و النموذج الرياضي يكتب بالشكل الآتي :

$$\log m_{ij} = \mu + \alpha_i^R + \beta_j^C + \gamma_{ij} \dots (3)$$

3-6-2 نموذج تأثير الصفوف Row Effect Model في بعض الأحيان تتطلب الحاجة الى معرفه او دراسة تأثيرات الصفوف لذلك قام الباحث (Goodman) باقتراح التحليل والنموذج لهذه الحالة وذلك باستخدام إعطاء الرتب للأعمدة والنموذج الرياضي موضح كالاتي:

$$\log m_{ij} = \mu + \alpha_i^R + \beta_j^C + j\phi_i \dots (4)$$

اذ ان:

$$\phi_i = \beta X_i$$

رتب الأعمدة بإعداد صحيحة.

4-6-2 نموذج تأثير الأعمدة Column Effect Model وفي حاله دراسة تأثير الأعمدة نقوم ترتيب الصفوف و النموذج الرياضي المستخدم لتحليل هذه التأثيرات هو كالاتي :

$$\log m_{ij} = \mu + \alpha_i^R + \beta_j^C + i\phi_j \dots (5)$$

اذ أن

$$\phi_j = \beta Y_j$$

وترتيب الصفوف بإعداد صحيحة.

5-6-2 نموذج Goodman RC Goodman RC Model والمسمى أيضا (Goodman I) قام الباحث (Goodman) بدراسة كلا التأثيرات للصفوف والأعمدة في وقت واحد ووضع النموذج المناسب لتحليل هذه التأثيرات كالاتي:

$$\log m_{ij} = \mu + \alpha_i^R + \beta_j^C + j\phi_i + i\phi_j \dots (6)$$

اذ ان (ϕ_j , ϕ_i) هي معلمات يتم تقديرها لمعرفة التأثيرات على العلاقة بين المتغيرين ويكون عددها (I+J) . ويتطلب هذا النموذج ان تكون مستويات الصفوف و الأعمدة ثابتة و يتم ترتيبها مسبقاً كما في النماذج السابقة (نموذج تأثير الصفوف , نموذج تأثير الأعمدة) ومن خلال هذا الشرط يتبين ان النموذج غير ملائم في أغلب الأحيان اي ان هناك تغير في ترتيب أصناف المتغيرات . لذلك اقترح النموذج الأكثر ملائمة من قبل الباحث (Goodman) وسماه بأنموذج (Goodman II) وكذلك يسمى بأنموذج (R.C) ويكتب كالاتي:

$$\log m_{ij} = \mu + \alpha_i^R + \beta_j^C + \beta \phi_i \phi_j \quad . . . (7)$$

وهذا الأنموذج يكون مفضلاً على النماذج الأخرى لكونه يسمح بتقدير معلمات الصفوف و الأعمدة حتى اذا كانت غير مرتبة او غير قابلة للترتيب .

Outliers in Contingency table

7-2 الشواذ في جداول التوافق [14] [15] [23] [24] [25]

تعد كثير من الدراسات الخلايا الشاذة على انها أخطاء بالتسجيل او بجمع البيانات و انها قد تحمل معلومات مهمة وعند إجراء التحليل بوجود هذه القيم تؤدي الى بعض النتائج التي تكون غير دقيقة فلا بد من الكشف عن هذه الخلايا ومعالجتها لذا فقد اوجدت عدة طرائق للكشف عن الخلايا الشاذة و التحليلات الإحصائية المستخدمة في تحليل بيانات لظاهرة ما تظهر وجود اختلاف في قيم المشاهدات وان هذا الاختلاف غالباً ما يكون ناتجة عن أخطاء في تسجيل جمع البيانات او عن طريق عوامل خارجية (عرضيه) تؤثر في التوزيع الطبيعي للظاهرة المدروسة .

وفي حاله التعامل مع الجداول التوافق فأنها سيتم التعامل مع الخلايا الشاذة الامر الذي يتطلب الكشف عنها بالاعتماد على حجم الخلية وفي مثل هذه الحالات يستوجب التعامل معها بحذر شديد .

بالإضافة الى ذلك ففي حاله وجود أكثر من خلية واحدة شاذة فأن مواقع تلك الخلايا في الجدول تكون حاسمه بالنسبة للكشف عنها لما لها من تأثير على طرائق تحليل البيانات .

ان تلك الحقيقة يمكن تمييزها من خلال مناقشه طرائق الكشف عن الشواذ و مناقشه نقاط الانهيار لجدول التوافق (2010) [41] وكذلك قدم (Fam . P, 2012) مصطلحات لأنماط الشواذ حيث توصل الى اختبارات حسن المطابقة عن طريق تطبيق تقنيات الإحصاءات الحيوية .

فقد اقتصرت البحوث للشواذ وفي جداول توافق ذات الاتجاهين بافترض أنموذج متعدد الحدود او أنموذج بواسون لوغاريتمي لذا سيتم التطرق في هذا البحث الى الخلايا الشاذة في جداول التوافق ثنائيه الإبعاد والتي تكون توزيعاتها متوافقة مع نماذج بواسون الخطية اللوغاريتمية والتي تجسد مفهوم او صورته القيم المتطرفة على انها خلايا شاذة .

الخلايا الشاذة في جداول التوافق ذو الاتجاهين تمت ملاحظتها من قبل (Barnett and Lemis) اذ تمت معالجتها مع أنموذج متعدد الحدود وذلك عن طريق توظيف المقدرات الحصينة وتحديد الاختبارات المناسبة ففي جداول التوافق يتم التعامل مع الخلية الشاذة بدلا من المشاهدة الشاذة حيث يكون مستند على التعامل مع كل خلية محسوبة لذا يتم التعامل معها بحذر لما لها من تأثير كبير على الأساليب الإحصائية لتحليل جداول التوافق بصوره عامة . الخلايا لشاذة مثالها كالمشاهدات التي تكون مبتعدة عن البيانات بنقطة معينه والتي تدعى منطقة الخلايا الشاذة

(α - outlier regions)(kuhnt).

لذلك فالمشاهدات المستقرة او الطبيعية في منطقة فضاء العينة تدعى (in liar) .

في أساليب التشخيص والتقدير الحصينة تكون الأوزان المصاحبة للمفردات كمؤشرات موزونة تعاقبية كمؤشرات للحالات الشاذة ، فبينما تبدو الأساليب الحصينة منافسة لأساليب التشخيص للنماذج إلا ان من غير الممكن ان تحل محلها طالما ان أساليب الانحدار الاعتيادية والمربعات الصغرى لا تزال واسعة الانتشار وان الحاجة الى أساليب التشخيص يبقى موجودا وان استخدام الطرق الحصينة لن يلغي وجود و أهمية التشخيص عموما رغم ان هناك إمكانية للاستغناء عن بعض تلك الأساليب .

8-2- طرائق تقدير معلمات الأنموذج اللوغاريتمي [2] [4] [12] [17] [26]

لتقدير معلمات أنموذج لوغاريتمي في حالة وجود خلايا شاذة او ملوثة تم الاعتماد على الطريقة الحصينة للتقدير اذ ان هذه الطريقة اقل حساسية إزاء القيم الشاذة .

وسوف نتطرق الى طريقة التقدير المستخدمة وهي :

أسلوب تقدير M .

— داله Huber .

— داله Tukeys .

— داله Andrews .

و إجراء المقارنة بين الازان المذكورة أعلاه , علماً ان الأنموذج المستخدم هو أنموذج لوغاريتمي بعد أدراج المتغيرات الوهمية (Dummy Variables) [8][14][16] وكالاتي:

$$\log m_{ij} = \lambda + \sum_{k=1}^{r-1} \lambda_k I_{ik} + \sum_{L=1}^{c-1} \lambda_L J_{Lj} \quad \dots (8)$$

$$i = 1,2, \dots, r ; J = 1,2, \dots, c$$

اذ ان المتغيرات الوهمية تأخذ القيم (1,0) :

$$\begin{cases} \text{if } I_{ii} = 1 \\ \text{o.w} = 0 \end{cases}$$

$$\begin{cases} \text{if } J_{jj} = 1 \\ \text{o.w} = 0 \end{cases}$$

وعلى سبيل المثال الأنموذج للخلية الأولى في جدول ثنائي (5 * 4) تكتب كما يلي:

$$\begin{aligned} \log m_{11} &= \lambda + \lambda_1 I_{11} + \lambda_2 I_{12} + \lambda_3 I_{13} + \lambda_1 J_{11} + \lambda_2 J_{12} + \lambda_3 J_{13} \\ &\quad + \lambda_4 J_{14} \quad \dots (9) \\ &= \lambda + \lambda_1 * (1) + \lambda_2 * (0) + \lambda_3 * (0) + \lambda_1 * (1) + \lambda_2 * (0) + \lambda_3 \\ &\quad * (0) + \lambda_4 * (0) \end{aligned}$$

$$\log m_{11} = \lambda + \lambda_1 + \lambda_1$$

تعريف المتغيرات الوهمية هو ذلك المتغير الذي يأخذ أرقاماً محددة من القيم لصفات مختلفة , و اهمية استخدام الرمز التأثري Effect Code للمتغيرات الوهمية للتمييز بين صفات المتغير بدلاً من الأرقام المتسلسلة التي تفقده الكثير من صفاته التحليلية.

1-8-2 الطريقة الحصينة Robust Method [4] [26]

لغرض تحسين نتائج التقديرات التي حصلنا عليها درسنا امكانية وجود الخلايا الشاذة ضمن جدول التوافق والتي تؤثر بشكل سلبي على نتائج التقدير. وعلى هذا الاساس استخدمنا اساليب او طرائق التقدير الحصينة لنموذج لوغاريتمي.

ومن الطرائق الحصينة المستخدمة للتقدير هي:

2-8-2 اسلوب تقدير M

تستند فكرة طريقة M على الاسلوب المتبع في طريقة المربعات الصغرى في تصغير مجموع مربعات الخطأ اذ ان طريقه المربعات الصغرى تتطلب تصغير المقدار الآتي :

$$\text{Min} \sum_{i=1}^n (n_{ij} - XB)^2 \quad \dots (10)$$

اما في طريقة تقدير - M فالمطلوب تصغير المقدار التالي:

$$\text{Min} \sum_{i=1}^n \rho (n_{ij} - XB) \quad \dots (11)$$

اذ ان ρ تمثل دالة محددة عادة Usually Convex Function ونشتقها جزئياً بالنسبة لموجه المعالم B وجعلها مساوية للصفر كالاتي:

$$\sum X_{ij} \psi (n_{ij} - XB) = 0 \quad \dots (12)$$

اذ ان ψ يمثل المشتقة الجزئية للدالة ρ بالنسبة للمعالم B اي أن:

$$\psi = \rho'$$

اذ ان المعادلة (12) تمثل مجموعة ρ من المعادلات غير خطيه وتحل باستخدام احدى الطرق العددية المعروفة للحصول على تقديرات M او باستخدام طريقه المربعات الصغرى الموزونة وفق الصيغة الآتية:

$$b = (X'AX)^{-1} X'AY \quad \dots (13)$$

اذ ان A تمثل مصفوفة الاوزان القطرية ويتم ايجاد الاوزان (عناصر القطر) على وفق الصيغة الآتية:

$$A_1 = \frac{\psi(n_{ij} - XB_0)}{(n_{ij} - XB_0)} \quad \dots (14)$$

في البداية يتم استخدام القيم الاولية B_0 لايجاد الاوزان ويمكن استخدام مقدرات المربعات الصغرى كقيم اولية فنجد قيم b_1 في التكرار الاول ثم تستخدم b_1 لايجاد الاوزان وبالتالي ايجاد b_2 في التكرار الثاني وهكذا نستمر حتى نحصل على قيم b متقاربة جداً في التكرارات , ولكي نحصل على مقدرات M تمتلك خاصية (Scale invariant) فان صيغة الدالة المطلوب تصغيرها ستكون كما يأتي :

$$\text{Min} \sum_{i=1}^n \rho \left(\frac{n_{ij} - XB}{s} \right) \quad \dots (15)$$

وبعد اشتقاق المعادلة (11) بالنسبة للموجة B ومساواتها للصفر نحصل على:

$$\sum_{i=1}^n X_{ij} \psi \left(\frac{n_{ij} - XB}{\hat{\sigma}} \right) = 0 \quad \dots (16)$$

اذ ان قيمة المقدر المعيار $\hat{\sigma}$ يقدر مرة واحدة فقط باستخدام القيم الاولية قبل بدء عملية التكرار من احد الصيغ الآتية :

$$\left. \begin{aligned} 1 - \hat{\sigma} &= 2.1 \text{ med}|r_i| \\ 2 - \hat{\sigma} &= \frac{1}{n-4} \sum_{i=3}^{n-2} |r_i| \\ 3 - \hat{\sigma} &= 1.48[\text{med}|r_i| - \text{med}r_i] \\ 4 - \hat{\sigma} &= [\text{med}|r_i|]/0.6745 \end{aligned} \right\} \quad \dots (17)$$

اذ ان r تمثل البواقي (Residuals) وان r_i تمثل البواقي المرتبة.

$$V(Y_{ij}) = m_{ij}$$

$$e_{ij} = (y_{ij} - \hat{m}_{ij}) / \sqrt{\hat{m}_{ij}}$$

وهناك عدة دوال مقترحة من دوال $\rho(\cdot)$ أو $\psi(\cdot)$ بحيث يكون المقدر الناتج عنها جيد ولا يتأثر بوجود الشواذ وفيما يأتي بعض الدوال والأوزان المقابلة لها:

1-2-8-2 دالة Huber [11]

$$\psi(x) = \begin{cases} K & \text{if } x \geq k \\ x & \text{if } |x| < k \\ -k & \text{if } x \leq -k \end{cases} \quad \dots (18)$$

2-2-8-2 دالة Tukey [11]

$$\psi(x) = \begin{cases} 0 & \text{if } |x| > k \\ x \left(1 - \left(\frac{x}{k}\right)^2\right) & \text{if } |x| \leq k \end{cases} \dots (19)$$

[11] Andrews دالة 3-2-8-2

$$\psi(x) = \begin{cases} \sin\left(\frac{x}{a}\right) & \text{if } -\pi a < x < \pi a \\ 0 & \text{o.w} \end{cases} \dots (20)$$

اذ ان

$$a = 1.142$$

$$k = 1.34$$

المبحث الثالث : الجانب التطبيقي

1-3 وصف البيانات [1] [6] [7] [8] [9] [10]

تعد البطالة احد أهم المشكلات الاقتصادية المستعصية في بلدان العالم كافة بما فيها البلدان المتطورة صناعيا ولكنها تتفاقم بشكل اكبر في البلدان النامية خصوصا العربية منها وذلك حسب دوريات منظمة العمل العربية والمنظمات العالمية اذ قدرت منظمة العمل العربية ان نسبة البطالة في الدول العربية تتجاوز 21% بما يساوي تقريبا 23 مليون عاطل عن العمل وذلك في عام 2003.

يعد الاقتصاد العراقي الغني بالموارد البشرية والمادية والذي يمكن ان يحدث طفرات نوعيه في مجال التنمية والتنمية المستدامة وذلك من خلال رفع مستوى الخدمات والإنتاجية , وبالرغم من كل تلك الثروات الطبيعية من النفط والغاز والثروات البشرية إلا انه نرى هناك قصورا واضحا أي وضع الخطط الاستراتيجية والحلول الناجمة لحل مشكلة البطالة لذا ينبغي على اصحاب الشأن حسن استخدام المواد المتاحة واستغلالها بالشكل الامثل وفقا لبرامج وخطط مدروسة كما ويجب التركيز على هذه المشكلة وإعطائها الاولويات وذلك بوصفها افة تهدد الاقتصاد العراقي ونسيجه الاجتماعي وخصوصا وان المجتمع العراقي يعاني من موجه شرسة تقودها قوى الارهاب العالمية والتي وجدت في الداخل بيئة خصبة سببها تدني الوضع الاقتصادي للفرد لذا كان من المناسب تناول هذه الظاهرة الخطرة بدراسة احصائية بغية تحليلها بأسلوب علمي يعزز الجوانب الوصفية في تحديد وتشخيص اثارها وتحليلها وفقا لمؤشرات يمكن عددها الاهم لمتخذي القرار وما اشارت اليه تقارير منظمات العمل الدولية فأن ابرز تلك المؤشرات هي التصنيف وفقا لمتغيري العمر والتحصيل الدراسي للفرد وهذا ما تم تناوله في سياق هذا البحث اذ تم جمعه من استمارات الاستبيان التي اعددها الجهاز المركزي للإحصاء في وزارة التخطيط الخاصة بمحافظة واسط اذ قام الباحث بجمع البيانات الخاصة بمتغيرات موضوع الدراسة (العمر, التحصيل الدراسي) لعينة من المجموعتين قوامها 1834 فردا.

2-3 تصنيف البيانات

طبقت عدة طرائق تقدير على جداول التوافق (6x5) وتم الحصول على تقديرات الجدول ذي الاتجاهين لبيانات البطالة المصنفة حسب الحالة التعليمية والفئات العمرية إذ صنف العمر الى خمس مجموعات (65-55, 55-45, 45-35, 35-25, 25-14) اما بالنسبة الى الحالة التعليمية فقد صنفت الى ست مجموعات (أمي , ابتدائي , متوسطة , اعدادية , دبلوم , بكالوريوس) وكان حجم العينة (1834) وتم

الحصول على البيانات من مديرية احصاء محافظة واسط حيث تم اخذ عينة من العاطلين عن العمل في المحافظة والجدول رقم (2) يبين البيانات التي تم استخدامها بهذا البحث .

الجدول (2) يمثل البيانات الحقيقية المصنفة حسب عمر الشخص وتحصيله الدراسي.

التحصيل الدراسي عمر الشخص	امي	ابتدائية	متوسطة	إعدادية	دبلوم	بكالوريوس	المجموع
14 – 25	134	446	78	46	42	66	812
25 – 35	86	118	76	60	22	92	454
35 – 45	30	82	60	34	64	38	308
45 – 55	46	34	22	26	10	16	154
55 - 65	42	22	14	16	6	6	106
المجموع	338	702	250	182	144	218	1834

3-3 تقدير نموذج الاقتران اللوغاريتمي الصفري

تم تقدير معاملات النموذج (2) باستعمال عدة طرائق التقدير وتمت المقارنة بين طرق التقدير باستعمال مقياس متوسط مربعات الخطأ (MSE) وطرق التقدير المستعملة هي:

تقدير معاملات نموذج الاقتران اللوغاريتمي الصفري باستعمال طريقة (M-estimate) الحصينة باستعمال دالة (Huber)

تم استعمال الصيغة المذكورة في معادلة (18) وباستخدام البرمجة بلغة (R) تم الحصول على النتائج التالية:

الجدول رقم (3) يبين نتائج طريقة (Huber)

m	α_1	α_2	α_3	α_4	β_1	β_2	β_3	β_4	β_5
2.1500872	.2047713	.0128633	1711721	1076649	4339371	.8309912	.4837549	.9446033	0.5894025

وقد كانت قيمة متوسط مربعات الخطأ (0.001149981)

تقدير معاملات نموذج الاقتران اللوغاريتمي الصفري باستعمال طريقة (M-estimate) الحصينة باستعمال دالة (Tukey)

تم استعمال الصيغة المذكورة في معادلة (19) وباستخدام البرمجة بلغة (R) تم الحصول على النتائج التالية:

الجدول رقم (4) يبين نتائج طريقة (Tukey)

m	α_1	α_2	α_3	α_4	β_1	β_2	β_3	β_4	β_5
2.1479584	0.2045686	1.011860	0.1710027	0.1075583	2.431527	1.829178	1.4822859	0.943668	0.5888189

وقد كانت قيمة متوسط مربعات الخطأ (0.001416919) تقدير معالم نموذج الاقتران اللوغاريتمي الصفري باستعمال طريقة (M-estimate) الحصينة باستعمال دالة (Andrews) تم استعمال الصيغة المذكورة في معادلة (20) وباستخدام البرمجة بلغة (R) تم الحصول على النتائج الآتية :

الجدول رقم (5) يبين نتائج طريقة (Andrews)

m	α_1	α_2	α_3	α_4	β_1	β_2	β_3	β_4	β_5
2.1522202	0.20497	1.01386	0.171342	0.10777	2.436351	1.832807	1.4852269	0.9455404	0.5899872

وقد كانت قيمة متوسط مربعات الخطأ (0.0009104305) .

الجدول (6) المعالم المقدرة وقيمة متوسط مربعات الخطأ لجميع طرائق التقدير للتجربة.

Methods Param.	Huber	Tukey	Andrews
m	2.1500872	2.1479584	2.1522202
α_1	0.2047713	0.2045686	0.2049745
α_2	1.0128633	1.0118605	1.0138681
α_3	0.1711721	0.1710027	0.1713420
α_4	0.1076649	0.1075583	0.1077717
β_1	2.4339371	2.4315272	1.8328076
β_2	1.8309912	1.8291783	2.4363517
β_3	1.4837549	1.4822859	1.4852269
β_4	0.9446033	0.9436680	0.9455404
β_5	0.5894025	0.5888189	0.5899872
MSE	0.00114998	0.00141691	0.000910435

وأظهرت النتائج ان اقل قيمة لمتوسط مربعات الخطأ كانت لدالة (Andrews) وبعدها جاءت دالة (Huber) وان افضل الطرق هي طريقة (Andrews) لذلك سوف تستخدم المعالم المقدرة بهذه الطريقة لإيجاد مقدرات الخلايا بالتعويض في الصيغة الآتية بغية الحصول على افضل مقدرات للخلايا :

$$e^{[\log m_{ij}]} = \text{common} + \text{row effect} + \text{Colum effect}$$

او

$$e^{[\log m_{ij}]} = m + \alpha_i + \beta_j \quad \dots (4 - 2)$$

الجدول (7) يمثل مقدرات الخلايا

صيل الدراسي عمر الشخص	أمي	ابتدائية	متوسطة	إعدادية	دبلوم	بكالوريوس
14 - 25	66	121	47	27	19	11
25 - 35	148	271	105	61	43	24
35 - 45	64	117	45	26	18	10

45 - 55	60	110	42	25	17	10
55 - 65	54	98	38	22	15	9

ان الهدف من هذه الدراسة هو إعطاء صورة واضحة عن ظاهرة البطالة وما لها من تداعيات تؤثر بصورة مباشرة وغير مباشرة على الواقع العراقي , بحيث تتمكن الجهات المختصة من التخطيط الجيد لتوفير فرص العمل.

ومن خلال التقديرات النهائية للخلايا بطريقة ((Andrews)) جاءت النتائج مطابقة الى حد ما لملامح البيانات الحقيقية والظاهرة اذ ان ما يمثل الثلث من العاطلين في البيانات الحقيقية هم من الأميين وحملة شهادة الابتدائية وهذا ما لوحظ ايضا بعد التقدير, كما وان اغلب العاطلين كانوا من الفئات العمرية الشابة ومن حملة الشهادات الدنيا. ويدل على كفاءة دالة ((Andrews)) بالنسبة لبقية الدوال المستعملة والتفوق الواضح لها على الطريقة التقليدية.

المبحث الرابع : الاستنتاجات و التوصيات

1 - الاستنتاجات

- 1- في الجانب التطبيقي وعلى بيانات واقعية كانت نتائج التطبيق العملي للطرائق الحصينة تؤكد بأن المقدرات الحصينة يمكن الاعتماد عليها في التطبيقات العملية والنظرية في حالة اختراق الافتراضات الأساسية.
- 2- التعامل مع الخلايا الشاذة في جداول التوافق (المتغيرات الفئوية) يكون مشابهاً إلى حد كبير إلى الشواذ في باقي المتغيرات الأخرى .
- 3- كما لوحظ من خلال حساب مقدرات الخلايا بأفضل طريقة تقدير في الجدول (7) ان اغلب العاطلين هم ممن تتراوح أعمارهم (14-35) أي فئات الشباب وان 38% من العاطلين هم من حملة الشهادة الابتدائية ولجميع الأعمار.
- 4- ان اغلب العاطلين هم من حملة الشهادات الدنيا ويكون مطابقاً للواقع وذلك لقلّة فرص العمل لديهم.

2- التوصيات

- استناداً الى النتائج التي توصلنا اليها من خلال الجانب التطبيقي لا بد من ذكر بعض التوصيات التي يرى الباحث بالإمكان الأخذ بها:
- 1- يمكن القول بان الطريقة الحصينة المعتمدة على دالة الوزن لـ (Andrews) هي من افضل الدوال في تقدير معلمات نموذج الاقتران .لذا نوصي باستعمال هذه الطريقة في التجارب المشابهة.
 - 2- يوصى بتطبيق طرائق التقدير الحصينة بالاعتماد على الدوال الوزنية المختلفة ولنماذج اقتران المختلفة التي تم ذكرها بالجانب النظري .
 - 3- توسعة التجربة والعمل على جداول توافق ذات رتب عليا أي (Three Ways, Four Ways, Multi Ways).
 - 4- قيام وزارة العمل والشؤون الاجتماعية وبالتنسيق مع الجهاز المركزي للإحصاء وتكنولوجيا المعلومات بخصر وتهيئة بيانات عن إعداد البطالة في العراق بهدف التعرف على حجمها ونوعها ومحتوياتها. لضمان استحداث برنامج وطني يقوم على التنسيق بين الوزارات المختلفة لخلق فرص العمل التي تتناسب مع مؤهلات العاطلين عن العمل , وبما يحقق العدالة الاجتماعية.
 - 5- القضاء على الأمية, و تطوير التعليم وربطه بالتدريب والتشغيل من خلال تعميق التعاون بين مؤسسات التعليم والتدريب والمؤسسات الإنتاجية والخدمية.
 - 6- دعم وتطوير القطاع الخاص بكافة أنشطته, والعمل على إزالة مخاوفه وإعطائه الضمانات لتوسيع نشاطه. بهدف خلق فرص العمل لاستيعاب الإعداد الفائضة عن حاجة سوق العمل.

7- وضع قانون لتخفيض سن التقاعد من 65 عام ال 55 عام وذلك بغية توفير وفسح المجال للطاقت الشابة من اخذ فرصها للعمل, ويستثنى من القانون (العقول والخبرات والقادة).

المصادر

- 1- احمد,مالك عبد الحسين (2011)"البطالة في العراق الأسباب والنتائج والمعالجة " الكلية التقنية والإدارية البصرة.
- 2- الجنابي, الشمري,عبود,ضوية سلمان حسن,نذير عباس إبراهيم,سهيل نجم,(2012):"تحليل البيانات المصنفة وتطبيقاتها",مكتبة الجزيرة للطباعة والنشر,العراق _بغداد.
- 3- الجنابي, مشتاق عبد الغني شخير (2002):"بعض تطبيقات النماذج اللوغارتمية الخطية في جداول توافق متعددة الاتجاهات " رسالة ماجستير في الرياضيات كليه تربيته جامعه المستنصرية .
- 4- الدباغ,ظافر عاصم مصطفى,(1999):"تحليل تباين حصين للنماذج الخطية"
- 5- اطروحة دكتوراة في الاحصاء,جامعة بغداد _كلية الادارة والاقتصاد.
- 6- القزاز, قتيبة نبيل نايف,(2001):" تحليل البيانات المصنفة في جداول التوافق باستخدام اسلوب بيز " رسالة ماجستير في الاحصاء,جامعة بغداد _كلية الادارة والاقتصاد.
- 7- القوي,محمد حسين عبد (2010) "البطالة المشكلة والعلاج"الاكاديمية الملكية للشرطة ,مملكة البحرين.
- 8- زكي,رمزي (1998) "الاقتصاد السياسي للبطالة " سلسلة كتب ثقافية شهرية يصدرها المجلس الوطني للثقافة والفنون والأدب - الكويت.
- 9- خلاصة نتائج مسح التشغيل والبطالة في العراق للفصل الثاني لعام 2008.
- 10- عبد الله , نداء حسين(2006):"واقع البطالة في العراق " بحث مقدم الى المعهد العربي للتدريب والبحوث الاحصائية للحصول على شهادة الدبلوم العالي في الاحصاءات التطبيقية.
- 11- كاظم,كامل علاوي (2011):"البطالة في العراق (الواقع,الأثار,آليات التوليد وسبل المعالجة)" بحث مقدم الى كلية الإدارة والاقتصاد جامعه الكوفة قسم الاقتصاد.
- 12- Aktas. S,(2002)," M-ESTIMATORS OF THE UNIFORM ASSOCIATION MODEL IN (R * C) CONTINGENCY TABLES", Hacettepe Journal of Mathematics and Statistics Volume 31, 83-88.
- 13- Agresti .A. (2002):" Categorical data analysis" Department of Statistics University of Florida Gainesville, Florida.
- 14- Bishop. Y. (2007):"Discrete Multivariate Analysis".
- 15- Daniela. F. and, Simonetta. C. (2008):" Outlier Identification Procedures for Contingency Tables in longitudinal data", ISTAT, Italy.
- 16- Fabio. R. (2011):" Outliers and Patterns of Outliers in Contingency Tables with Algebraic Statistics" department distal, University del piedmont Orientale viable Teresa Michel, 11 Alessandria, Italy.
- 17- Fam .P.(2012):" Analysis of Two-way Contingency Tables" Applied Mathematical Sciences, Vol. 6, No. 79,PP. 3917 – 3925 .
- 18- Fox .J. (2002):" Robust Regression" Appendix to An R and S-PLUS Companion to Applied Regression.
- 19- Goodman .L. (1979):"Simple models for the analysis of association in cross-classifications having ordered categories" JASA, Vol. 74, NO. 367, PP. 537-552.

- 20- Goodman .L. (1981):" Association Models and the Bivariate Normal for Contingency Tables with Ordered Categories", *Biometrika*, Vol. 68, No. 2 (Aug., 1981), pp. 347-355.
- 21- Goodman .L. (1981):" Association models and canonical correlation in the analysis of cross-classifications having ordered categories" *JASA*, Vol. 76, NO. 374, PP. 320-334.
- 22- Goodman .L. (1981):" Three elementary views of Log linear models for the analysis of cross-classifications having ordered categories" *JASA*, Vol. 12, PP. 193-239.
- 23- Goodman .L.(1985):" The Analysis of Cross-Classified Data Having Ordered and/or Unordered Categories:Association Models, Correlation Models, and Asymmetry Models for Contingency Tables With or Without Missing Entries", *The Annals of Statistics*, Vol. 13, No. 1 , pp. 10-69.
- 24- Hubert .M. (1997):"The breakdown value of the L_1 estimator in contingency tables" *statistics & probability letters*, Vol.33, pp.419-425.
- 25- Kuhnt. S. (2010):"Breakdown concepts for contingency tables" *Published online*, Vol. 13, pp281-294.
- 26- Kuhnt. S.; Rapallo. F and Rehage .A. (2012):"Outlier Detection in Contingency Tables based on Minimal Patterns".
- 27- Shane .K and Simonoff .J (2001):"A Robust Approach to Categorical data analysis" *Journal of Computational and Graphical Statistics*, Vol. 10, NO.1, pp.135-157.
- 28- Williams .D and James .E. (1972):'Analysis of contingency tables having ordered response categories" *JASA*, Vol. 67, NO. 337, PP. 55-63.
- 29- Xie .Y. (2000):" Association Model" *Institute for Social Research*, 426 Thompson Street, University of Michigan, Ann.